PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-132455
(43)Date of publication of application : 10.05.2002

(51)Int.Cl. G06F 3/06
G06F 12/00
G06F 13/10

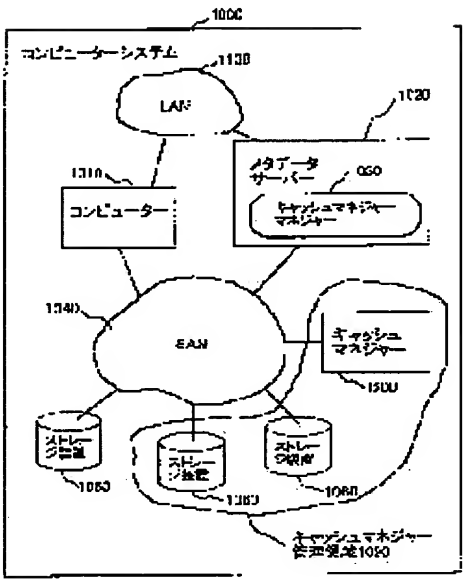
(21)Application number : 2000-324868
(22)Date of filing : 25.10.2000

(71)Applicant : HITACHI LTD
(72)Inventor : ACHIWA KIYOUSUKE
SATO TAKAO

(54) CACHE MANAGER AND COMPUTER SYSTEM INCLUDING IT

(57)Abstract:
PROBLEM TO BE SOLVED: To provide a storage device that is connected to a SNA and has not a cache function with a cache function to accelerate access to the storage device.
SOLUTION: A cache manager 1200 provides the storage device 1060 without cache function with the cache function. A computer 1010 issues an input/output request for accessing the data stored in the storage device 1060 to the cache manager 1200. The cache manager 1200 converts the positional information on the data received together with the input/output request into the address of the target storage device 1060 and its data positional information, and provides the storage device 1060 with the cache function.

図 :



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（Ａ）**　(11)特許出願公開番号
特開2002－132455
（P2002－132455A）
(43)公開日　平成14年５月10日(2002.5.10)

(51)Int.Cl.⁷	識別記号	ＦＩ	テーマコード（参考）
Ｇ０６Ｆ　3/06	３０２	Ｇ０６Ｆ　3/06	３０２Ａ　５Ｂ０１４
12/00	５１４	12/00	５１４Ｍ　５Ｂ０６５
	５４５		５４５Ａ　５Ｂ０８２
13/10	３４０	13/10	３４０Ａ

審査請求　未請求　請求項の数10　ＯＬ　（全 13 頁）

(21)出願番号　特願2000－324868(P2000－324868)

(22)出願日　平成12年10月25日(2000.10.25)

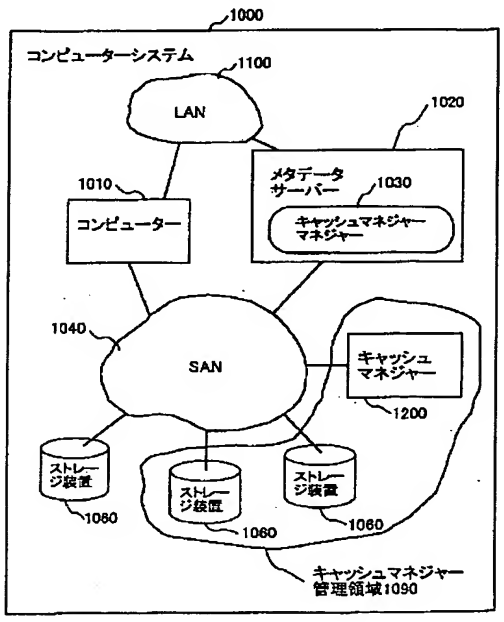
(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地
(72)発明者　阿知和　恭介
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内
(72)発明者　佐藤　孝夫
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内
(74)代理人　100068504
弁理士　小川　勝男　（外２名）

Ｆターム(参考)　5B014 EB05
5B065 BA01 CH01
5B082 FA02 HA00

(54)【発明の名称】　キャッシュマネジャー及びこれを含むコンピュータシステム

(57)【要約】
【課題】　SANにつながったキャッシュ機能を持たない記憶装置にキャッシュ機能を提供し、その記憶装置へのアクセスを高速化する。
【解決手段】　キャッシュマネジャー1200は、キャッシュ機能をもたないストレージ装置1060に対してキャッシュ機能を提供する。コンピュータ1010は、ストレージ装置1060に格納されているデータにアクセスするための入出力要求をキャッシュマネジャー1200に対して発行する。キャッシュマネジャー1200は、入出力要求とともに受け取ったデータの位置情報を目的とするストレージ装置1060のアドレスとそのデータ位置情報に変換し、そのストレージ装置1060に対するキャッシュ機能を提供する。

図 1



【特許請求の範囲】
【請求項１】複数の記憶装置と、前記記憶装置に格納されるデータにアクセスするコンピュータと、前記記憶装置の少なくとも１台にキャッシュ機能を提供する装置であるキャッシュマネジャーと、前記記憶装置、前記コンピュータ及び前記キャッシュマネジャー間を接続するネットワークとを有するコンピューターシステムであって、前記コンピュータは、前記記憶装置に格納されているデータにアクセスするための入出力要求を前記キャッシュマネジャーに対して発行することを特徴とするコンピューターシステム。
【請求項２】前記コンピューターシステムは、前記記憶装置のネットワークアドレスと、データを格納するブロック番号との代りに、前記キャッシュマネジャーのネットワークアドレスと、前記記憶装置のネットワークアドレス及び前記ブロック番号を組み込んだブロック番号とを前記コンピュータに通知する手段を有することを特徴とする請求項１記載のコンピュータシステム。
【請求項３】前記通知する手段をメタデータサーバ内に設けることを特徴とする請求項２記載のコンピューターシステム。
【請求項４】前記コンピュータシステムは、前記キャッシュマネジャーのネットワークアドレスと割り当てる前記記憶装置のネットワークアドレスとの対応を格納する記憶手段と、外部からの指示に従って前記対応を変更する手段を有することを特徴とする請求項１項記載のコンピューターシステム。
【請求項５】前記記憶手段と前記対応を変更する手段をメタデータサーバ内に設けることを特徴とする請求項４記載のコンピューターシステム。
【請求項６】前記コンピュータシステムは、前記キャッシュマネジャーの識別子と割り当てる前記記憶装置の識別子との対応を格納する記憶手段と、あらかじめ設定された前記記憶装置の最大数に達するまで前記記憶手段に前記キャッシュマネジャーの識別子と対応する前記記憶装置の識別子を設定する手段とを有することを特徴とする請求項１項記載のコンピューターシステム。
【請求項７】前記記憶手段は、さらに前記記憶装置に対応してそのアクセス回数を格納し、前記アクセス回数の高い記憶装置から順に前記最大数に達するまでキャッシュ機能を提供する記憶装置を選択する手段を有することを特徴とする請求項６記載のコンピューターシステム。
【請求項８】複数の記憶装置と、前記記憶装置に格納されるデータにアクセスするコンピュータと、キャッシュマネジャーとがネットワークを介して接続されるコンピューターシステム中の装置であるキャッシュマネジャーであって、前記キャッシュマネジャーは、前記コンピュータから受け取った当該キャッシュマネジャーのネットワークアドレスとデータの位置情報を目的とする前記記憶装置のネットワークアドレスと当該記憶装置内のデータの位置情報に変換する手段と、前記の記憶装置にキャッシュ機能を提供する手段とを有することを特徴とするキャッシュマネジャー。
【請求項９】複数の記憶装置と、前記記憶装置に格納されるデータにアクセスするコンピュータと、キャッシュマネジャーとがネットワークを介して接続されるコンピューターシステム中の装置であるキャッシュマネジャーであって、前記キャッシュマネジャーは、当該キャッシュマネジャー宛てに送られたデータの位置情報を目的とする前記記憶装置上のデータの位置情報に変換する手段と、前記コンピュータとの間に転送されるデータを一時保存するキャッシュ手段とを有することを特徴とするキャッシュマネジャー。
【請求項１０】複数の記憶装置と、前記記憶装置に格納されるデータにアクセスするコンピュータと、キャッシュマネジャーとがネットワークを介して接続されるコンピューターシステム中の装置であるキャッシュマネジャーであって、前記キャッシュマネジャーは、当該キャッシュマネジャー宛てに送られたデータの位置情報を目的とする前記記憶装置上のデータの位置情報に変換する手段と、前記コンピュータと前記記憶装置との間に転送されるデータを一時保存するキャッシュ手段と、前記キャッシュマネジャーの識別子と割り当てる前記記憶装置の識別子との対応を格納する記憶手段と、あらかじめ設定された前記記憶装置の最大数に達するまで前記記憶手段に前記キャッシュマネジャーの識別子と対応する前記記憶装置の識別子を設定する手段とを有することを特徴とするキャッシュマネジャー。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は、ネットワーク上でキャッシュ機能を持たない記憶装置に対してキャッシュ機能を提供するキャッシュマネジャー及びこれを含むコンピュータシステム関する。
【０００２】
【従来の技術】ストレージエリアネットワーク（以下SANと記述）は、コンピューターネットワークの一種であり、主としてストレージトラフィックが流れる。SANはストレージデバイスやストレージサブシステムと、それらに対してデータの読み書きを行う一台以上のコンピューターなどから構成され、これらの装置間はファイバーチャネルで結ばれ、デバイス同士の間ではFCPと呼ばれるファイバーチャネル上のSCSIプロトコルを使ってデータのやりとりをすることが多い。SANに関する記述の一例は「Building Storage Networks」Marc Farley著、Osborne出版、28-29ページにある。
【０００３】SANにつながっている複数のコンピュータ一間でSAN上のデータを共有する方法の一つとして、前記「Building Storage Networks」475-476ページにあるように、共有ファイルシステムとファイルに関する管理

情報を扱うメタデーターサーバー（「Building Storage Networks」では「Connection Broker」と記述）を使う方法がある。SAN上のストレージ装置にアクセスしようとするコンピューターは、メタデータサーバーに対してLAN経由でファイルの物理的な格納位置を問い合わせ、メタデータサーバーは問い合わせのあったファイルの物理的な位置情報をやはりLAN経由で返す。位置情報はストレージ装置番号とストレージ装置内の論理ブロック番号と論理ブロック長およびそれらのリストなどある。そして、そのコンピューターは受け取った位置情報の示すストレージ装置に対し、SAN経由で位置情報の示す論理ブロック番号をアクセスする要求を発行する。その後のデータ転送もSAN経由で行われる。

【０００４】SCSIには、デバイスの構成情報や属性を問い合わせる目的のInquiryコマンドがある。デバイスから返されるInquiryデータには、デバイスタイプコード、ベンダーID、プロダクトID等の他、ベンダー固有情報を入れることもできる。SCSIのInquiryコマンドについては、「SCSI-2詳細解説」菅谷誠一著、CQ出版社、127-129ページに記載がある。

【０００５】SCSIには、デバイスに対してパラメータを設定するMode Selectコマンドがある。Mode Selectコマンドを使ってベンダーユニークなパラメータを設定することができる。SCSIのMode Selectコマンドについては、「SCSI-2詳細解説」菅谷誠一著、CQ出版社、141-144ページに記載がある。

【０００６】

【発明が解決しようとする課題】上記従来技術のSANに接続されるストレージには、キャッシュ機能を有するものと、キャッシュ機能をもたないものとがある。キャッシュ機能をもたないストレージに対するデータの読み書き性能は、キャッシュ機能をもつストレージに比べて低下するので、キャッシュ機能をもたないストレージに対してもキャッシュ機能を提供するのが望ましい。

【０００７】またキャッシュ機能をもたないストレージに対するコンピュータからのアクセス頻度の変動に対し、アクセス頻度が高いときに動的にキャッシュ機能を提供できるのが望ましい。

【０００８】本発明の第１の目的は、ネットワークに接続されるキャッシュ機能をもたない記憶装置に対して、コンピューター側の設定を変更せずにキャッシュ機能を提供することにある。

【０００９】本発明の第２の目的は、キャッシュ機能を提供する記憶装置を動的に選択可能にすることにある。

【００１０】本発明の第３の目的は、上記のキャッシュ機能を提供する装置をキャッシュマネジャーとして上記ネットワークに接続したときに、キャッシュマネジャーによって増加するネットワークトラフィックを制限することにある。

【００１１】

【課題を解決するための手段】本発明は、複数の記憶装置と、この記憶装置に格納されるデータにアクセスするコンピュータと、この記憶装置の少なくとも１台にキャッシュ機能を提供する装置であるキャッシュマネジャーと、記憶装置、コンピュータ及びキャッシュマネジャー間を接続するネットワークとを有するコンピューターシステムであって、コンピュータは、記憶装置に格納されているデータにアクセスするための入出力要求をキャッシュマネジャーに対して発行するコンピューターシステムを特徴とする。

【００１２】また本発明のコンピュータシステムは、キャッシュマネジャーの識別子と割り当てる記憶装置の識別子との対応を格納する記憶手段と、あらかじめ設定された記憶装置の最大数に達するまでこの記憶手段にキャッシュマネジャーの識別子と対応する記憶装置の識別子を設定する手段とを有するコンピューターシステムを特徴とする。

【００１３】さらに本発明は、さらに上記の記憶手段が上記の記憶装置に対応してそのアクセス回数を格納し、アクセス回数の高い記憶装置から順に最大数に達するまでキャッシュ機能を提供する記憶装置を選択する手段を有するコンピューターシステムを特徴とする。

【００１４】

【発明の実施の形態】（１）第１の実施形態

図１は、本発明を適用するコンピューターシステム1000を示す。コンピューターシステム1000は、SAN1040、データを格納するストレージ装置1060、データを利用するコンピューター1010、後述するキャッシュマネジャー1200、キャッシュマネジャー1200を管理するキャッシュマネジャーマネジャー1030がその上で動き、コンピューター1010からのファイルの位置情報問い合わせに対し、ファイルの位置情報を返す働きをするメタデータサーバー1020、コンピューター1010とメタデータサーバー1020を結び、ファイルの位置情報の問い合わせ、返答が流れるLAN1100から構成される。キャッシュマネジャーマネジャー1030は、後述するドライブ管理テーブル1400、キャッシュマネジャー管理テーブル1800、ドライブソートテーブル3200およびカレントソート番号3220を管理する。

【００１５】キャッシュマネジャー1200がキャッシュ機能を提供するストレージ装置1060、及びキャッシュマネジャー1200をあわせてキャッシュマネジャー管理領域1090と呼ぶ。

【００１６】本実施形態では、SAN1040につながっている全てのデバイスは、そのネットワークアドレスとして、固有のSANアドレスを持つ。SAN1040上のデバイスが他のデバイスにアクセスする時には、このSANアドレスで相手を指定する。

【００１７】図２は、複数台のコンピューター1010、複数台のキャッシュマネジャー1200が存在するコンピューターシステムの構成例であるコンピューターシステムA1

001を示す。コンピューターシステムA1001において、2台のキャッシュマネジャー1200が存在する。キャッシュマネジャーA1201はドライブB1062、ドライブC1063の2台のストレージ装置1060に対してキャッシュ機能を提供し、キャッシュマネジャー管理領域A1091を構成する。キャッシュマネジャーB1202は、ストレージ装置1060に属するドライブA1061に対してキャッシュ機能を提供し、キャッシュマネジャー管理領域B1092を構成する。ストレージ装置1060であるドライブD1064はどのキャッシュマネジャー1200からもキャッシュ機能の提供を受けていない。

【００１８】キャッシュ機能を提供されないドライブD1064に対するコンピューター1010からのアクセス要求は、従来の技術で説明したのと同様に以下のように行われる。まず、コンピューター1010はLAN1100経由でメタデータサーバー1020に対し、アクセスしたいファイルの位置情報を問い合わせる。メタデータサーバー1020は、ファイルのデータがドライブD1064上のある場所に格納されていることを突き止め、その位置情報をLAN1100経由でコンピューター1010に返す。コンピューター1010は受け取った位置情報を元に、ドライブD1064に対し、受け取った位置情報の示す格納データをアクセスするようSAN1040経由で要求する。このように、本実施形態において、コンピューター1010がファイルをアクセスする際には、まず位置情報をメタデータサーバー1020に問い合わせ、その位置情報を元にストレージ装置1060にアクセスする手順をとる。

【００１９】図３は、キャッシュマネジャー1200の構成を示す。キャッシュマネジャー1200は、SAN1040とのデータやりとりを行うインターフェース1210、データを転送するDMA1220、キャッシュ管理やデータ転送指示を行うCPU1240、CPU1240が実行するプログラムを格納するROM1230、コンピュータ1010とストレージ装置1060との間に転送されるデータを一時保存するキャッシュメモリとして機能し、CPUが変数を保持するのにも使うRAM1250から構成される情報処理装置である。RAM1250上には後述するキャッシュ管理テーブル1600、及び後述するカレントキャッシュブロック番号1700、更に後述するキャッシュドライブ管理テーブル3800が格納される。

【００２０】図４は、メタデータサーバー1020上にあり、キャッシュマネジャーマネジャー1030が管理するドライブ管理テーブル1400を示す。ドライブ管理テーブル1400はストレージ装置1060を管理するテーブルであり、ストレージ装置1060一台につき一つのドライブ管理テーブルエントリ1420を持つ。各エントリに付けられた番号は、SAN内でのストレージ装置1060のドライブ番号を示す。ドライブ管理テーブルエントリ1420は、アクセス回数1430とキャッシュマネジャーSANアドレス1440及びそのストレージ装置1060のSANアドレス1450の情報を持つ。アクセス回数1430は、当該ストレージ装置1060のアクセス回数を計数するカウンタである。キャッシュマネジャーSANアドレス1440は、そのストレージ装置1060にキャッシュ機能を提供するキャッシュマネジャー1200が存在すればそのキャッシュマネジャー1200のSANアドレスが入り、対応しているストレージ装置1060がキャッシュ機能を有している場合は値-1が入り、対応しているストレージ装置1060がキャッシュ機能を有しておらず、かつキャッシュ機能を提供しているキャッシュマネジャー1200がなければ値-2が入る。なおストレージ装置1060がキャッシュ機能を有しているか否かの区分は、通常キャッシュ機能と呼ばれる機能を有しているか否かの程度の区分でよい。SANアドレス1450は、当該ストレージ装置1060のSANアドレスを格納する。エントリに対応しているストレージ装置1060が存在しない場合には、SANアドレス1450に無効を意味する値-1が入る。

【００２１】図５は、メタデータサーバー1020上にあり、キャッシュマネジャーマネジャー1030が管理するキャッシュマネジャー管理テーブル1800である。そのエントリには、対応するキャッシュマネジャー1200がキャッシュ機能を提供できるストレージ装置1060の最大数を示す対応可能ドライブ台数1820の情報と、当該キャッシュマネジャー1200のSANアドレス1830が入る。対応するキャッシュマネジャー1200が存在しないエントリのSANアドレス1830には無効を示す値-1が入る。

【００２２】図６は、メタデータサーバー1020上にあり、キャッシュマネジャーマネジャー1030が管理するドライブソートテーブル3200とカレントソート番号3220を示す。ドライブソートテーブル3200はそのエントリにストレージ装置1060のドライブ管理テーブル1400のエントリ番号（ストレージ装置1060の番号）が入り、後述するキャッシュマネジャー構成変更処理3000で使用される。カレントソート番号3220にはドライブソートテーブル3200のエントリ番号が入り、同様にキャッシュマネジャー構成変更処理3000で使用される。

【００２３】ここで、本実施形態では単純化のために、全てのファイルが同じ長さを持ち、それはストレージ装置1060のブロック長と同じであるとする。このことは発明の本質とは直接関係のない単純化であり、ファイルごとにの長さが異なる場合においても本発明を適用可能であることは言うまでもない。

【００２４】図７は、キャッシュマネジャー1200上にあり、キャッシュデータを管理するキャッシュ管理テーブル1600及び次にリプレースされるキャッシュブロックの番号を示すカレントキャッシュブロック番号1700を示す。キャッシュ管理テーブル1600は、キャッシュデータを格納するキャッシュブロックごとにキャッシュ管理テーブルエントリ1620を持つ。キャッシュ管理テーブルエントリ1620は、そのキャッシュブロックに入っているデータに対応するストレージ装置1060のSANアドレス1630、ストレージ装置1060内の対応するブロック番号であ

るLBA(Logical Block Address)1640、当該キャッシュブロックに格納されているデータがストレージ装置1060に未反映であることを示すダーティーフラグ1650の情報を持つ。有効なデータを保持していないキャッシュブロックに対応するエントリのSANアドレス1630には無効を意味する値-1が入る。

【００２５】本実施形態では、ファイル長さとストレージ装置1060のブロック長を同じとしたと述べたが、キャッシュブロックのサイズもこれらと同じ長さとする。そのため、キャッシュブロック一つに一つのファイルが入る。

【００２６】ディスクキャッシュのリプレースにはLRU(Least Recently Used)等のアルゴリズムを利用することが一般的であるが、本実施形態では単純化のために以下のようにした。つまり、カレントキャッシュブロック番号1700が次にリプレースされるキャッシュブロックを指すようにし、それが指すキャッシュブロックがリプレースされて新たなデータが入ったら、カレントキャッシュブロックを一つ進める。つまり、カレントキャッシュブロック番号1700に1を加え、カレントキャッシュブロック番号1700がそのキャッシュマネジャー1200の持つキャッシュブロック数と同じになったらカレントキャッシュブロック番号1700を0に戻す。

【００２７】図８は、キャッシュマネジャー1200上にあり、当該キャッシュマネジャー1200がキャッシュ機能を提供するストレージ装置1060を管理するキャッシュドライブ管理テーブル3800である。このテーブルは当該キャッシュマネジャー1200が管理可能なストレージ装置1060の数と同じだけのエントリを持ち、エントリにはキャッシュ機能を提供するストレージ装置1060のSANアドレス3820が入る。対応するストレージ装置1060の存在しないエントリのSANアドレス3820には無効を意味する値-1が入る。

【００２８】図９は、コンピューター1010がSAN1040につながっているストレージ装置1060に対してアクセスする際のホストI/O処理2000のフローチャートを示す。

【００２９】まずステップ2010において、コンピューター1010はメタデータサーバー1020に対し、アクセスしたいファイルのパスを指定してそのファイルの格納位置を問い合わせる。次にステップ2020で、メタデータサーバー1020よりそのファイルの格納されているストレージ装置1060のSANアドレスと装置内の格納ブロック番号からなる格納位置情報を受け取り、ステップ2030で受領した位置情報の示すストレージ装置1060に対し、格納ブロック番号のデータをアクセスするようI/O要求を発行し、ステップ2040で目的とするデータの送受信を行ってホストI/O処理2000は終了する。

【００３０】図１０は、ホストI/O処理2000のステップ2010でコンピューター1010からファイルの格納位置を問い合わされた時に、メタデーターサーバー1020が実行するメタデータサーバー問い合わせ処理2200のフローチャートである。

【００３１】まずステップ2210で、コンピューター1010がアクセスしたいファイルのパスを受領する。そしてステップ2220でそのパスの示すファイルを格納するストレージ装置1060のSANアドレスとその装置内のブロック番号からなる位置情報を得て、ステップ2230でアクセス対象のストレージ装置1060に対応するドライブ管理テーブルエントリ1420のアクセス回数1430に1を加える。そして、同エントリ1420のキャッシュマネジャーSANアドレス1440を調べて、対象ストレージ装置1060にキャッシュ機能を提供しているキャッシュマネジャー1200があるかどうかの判定を行う。あると判定された場合には、ステップ2250で、以下の式を用いて位置情報の変換を行う。

【００３２】ブロック番号＝対象ストレージ装置のSANアドレス×1000000＋対象ブロック番号SANアドレス＝キャッシュマネジャーの持つSANアドレスこの変換により、アクセス対象となるSANアドレスはキャッシュマネジャー1200のSANアドレスとなり、ブロック番号には元々のアクセス対象のストレージ装置1060のSANアドレスとそのストレージ装置1060内ブロック番号を組み合わせた値が入ることになる。なおここでは、LBAの最大値は１０進換算で６桁以下で表現される数値を仮定している。1000000の値は例であり、LBAの最大桁数を越える桁数の数値であればよい。このブロック番号は、対象ストレージ装置のSANアドレスと対象ブロック番号が、後で両者を分離できるように組み込まれていればよい。そしてステップ2260において、得られたSANアドレスとブロック番号からなる位置情報をコンピューター1010に返し、メタデータ問い合わせ処理2200は終了する。

【００３３】一方、ステップ2240で無いと判定された場合には、ステップ2260にジャンプする。

【００３４】なお、ステップ2230、ステップ2240およびステップ2250の処理はキャッシュマネジャーマネジャー1030が行う。

【００３５】図１１は、コンピューター1010がキャッシュマネジャー1200に対してリードコマンドを発行したときに、キャッシュマネジャー1200で行われるキャッシュマネジャーリード処理2400のフローチャートである。

【００３６】まずステップ2410で、コンピューター1010からリード要求を受領し、ステップ2420で位置情報の変換を行う。位置情報の変換は以下の式で行う。

【００３７】SANアドレス＝ブロック番号÷1000000（小数点以下切り捨て）

ブロック番号＝ブロック番号を1000000で割った余り

この変換はメタデータサーバー問い合わせ処理2200のステップ2250で行った変換の逆変換であり、コンピューター1010が要求したファイルが格納されているストレージ装置1060のSANアドレスとブロック番号が得られる。キャッシュマネジャー1200は、キャッシュドライブ管理テ

ーブル3800を参照して得られたSANアドレスがキャッシュドライブ管理テーブル3800に登録された正しいアドレスか否かをチェック可能である。次にステップ2430で、キャッシュ管理テーブル1600を調べ、ステップ2420で得られたSANアドレスとブロック番号がそれぞれSANアドレス1630とLBA1640と一致するキャッシュ管理テーブルエントリ1620があるかどうか（つまり要求データがキャッシュ上にあるかどうか）調べる。一致せずに無いと判定された場合には、ステップ2440で、ステップ2420で得られたストレージ装置1060に対し、得られたブロック番号のブロックを読むようリード要求を発行する。そしてステップ2450でカレントキャッシュブロック番号1700が示すキャッシュブロックにデータを読み込み、ステップ2455でカレントキャッシュブロックを一つ進める（カレントキャッシュブロックの進め方についてはすでに説明した通りである）。そして、ステップ2460で、データを読み込んだキャッシュブロックに対応するキャッシュ管理テーブルエントリ1620を設定する。具体的には、SANアドレス1630にはステップ2420で求めたSANアドレスを入れ、同様にLBA1640にはステップ2420で求めたブロック番号を入れ、ダーティーフラグ1650は下げておく（リセットする）。そしてステップ2470において、キャッシュ上の要求データをコンピューター1010に転送してキャッシュマネジャーリード処理2400は終了する。

【００３８】ステップ2430でキャッシュ上にあると判定された場合には、ステップ2470にジャンプする。

【００３９】図１２は、コンピューター1010がキャッシュマネジャー1200に対してライトコマンドを発行したときに、キャッシュマネジャー1200で行われるキャッシュマネジャーライト処理2600のフローチャートである。

【００４０】まずステップ2610で、コンピューター1010からライト要求を受領し、ステップ2620で位置情報の変換を行う。位置情報の変換はキャッシュマネジャーリード処理2400のステップ2420で行われる処理と同様にして以下の式で行う。

【００４１】SANアドレス＝ブロック番号÷1000000（小数点以下切り捨て）

ブロック番号＝ブロック番号を1000000で割った余り

次にステップ2630でキャッシュ管理テーブル1600を調べ、ステップ2620で得られたSANアドレスとブロック番号がSANアドレス1630とLBA1640と一致するキャッシュ管理テーブルエントリ1620があるかどうか（つまり要求データがキャッシュ上にあるかどうか）調べる。一致せずに無いと判定された場合には、ステップ2640でカレントキャッシュブロック番号1700の示すキャッシュブロックを受領位置として設定し、ステップ2650でカレントキャッシュブロックを進める。そしてステップ2670で受領位置のキャッシュブロックにコンピューター1010からのライトデータを受領し、ステップ2680で受領位置のキャッシュブロックに対応するキャッシュ管理テーブルエントリ1620を書き換える。具体的には、SANアドレス1630にはステップ2620で求めたSANアドレスを入れ、LBA1640にも同様にステップ2620で求めたブロック番号を入れる。そしてステップ2685で同じキャッシュ管理テーブルエントリ1620のダーティーフラグ1650を上げ（セットし）、ステップ2690でコンピューター1010に対してライトコマンドの完了報告を返す。その後ステップ2700で、ステップ2620で求めたストレージ装置1060に対し、同じようにステップ2620で求めたブロック番号に対するライトコマンドを発行し、ステップ2710で、コンピューター1010から受け取ったデータをステップ2620で求めたストレージ装置1060に対して転送する。そしてステップ2720で、ステップ2685で上げたダーティーフラグ1650を下げて（リセットし）、キャッシュマネジャーライト処理2600は終了する。

【００４２】ステップ2630で目的のデータがキャッシュ上にあると判定された場合には、ステップ2660で、SANアドレスとブロック番号がSANアドレス1630とLBA1640と一致したキャッシュ管理テーブルエントリ1620を受領位置として、ステップ2670にジャンプする。

【００４３】本実施形態においては、ストレージ装置1060に未反映のキャッシュデータをストレージ装置1060に反映する処理を、キャッシュマネジャーライト処理2600において行っているが、ライト処理ごとに行わず、30秒に一度まとめて未反映データを反映するなど、ある程度時間をおいてから行うことも可能である。このような場合においても本発明を適用可能であることは言うまでもない。

【００４４】図１３は、SAN1040にストレージ装置1060やキャッシュマネジャー1200等のデバイスが新たにつながったときに、メタデータサーバー1020で実行されるデバイス認識処理2800のフローチャートを示す。

【００４５】まずステップ2810でデバイスがSAN1040に追加されたことを認識する。メタデータサーバー1020がデバイスの追加を認識する方法はいくつかあるが、本実施例ではSAN1040の管理者が、追加されたデバイスのSANアドレスをメタデータサーバ1020に入力するものとする。次にステップ2820で、メタデータサーバー1020は、追加されたデバイスに対し、デバイス種別確認コマンドを発行する。このコマンドはSCSIにあるInquiryコマンドを用い、デバイス種別等を問い合わせる。そしてステップ2830で、新規追加デバイスからの、デバイス種別確認コマンドに対するレスポンスを受領する。このレスポンスには、そのデバイスがキャッシュマネジャー1200かあるいはそれ以外のストレージ装置1060か、さらにはそれ以外のデバイスかがわかる情報と、キャッシュマネジャー1200であれば、キャッシュ機能を提供できるストレージ装置1060の台数情報、キャッシュマネジャー1200以外のストレージ装置1060であれば当該ストレージ装置1060がキャッシュ機能を有するかどうかの情報が入ってい

るものとする。次にステップ2840で、受領したレスポンスを見て新規追加されたデバイスがストレージ装置1060かどうかを調べ、そうでなければステップ2850で新規追加されたデバイスがキャッシュマネジャーかどうかを調べる。キャッシュマネジャー1200と判定された場合には、ステップ2870でキャッシュマネジャー管理テーブル1800の空きエントリに対応可能ドライブ台数1820としてキャッシュ機能を提供できるストレージ装置1060の台数を代入し、更にそのキャッシュマネジャー1200のSANアドレスをSANアドレス1830に代入して、次にステップ3000で後述するキャッシュマネジャー構成変更処理3000を行ってデバイス認識処理2800は終了する。なお、キャッシュマネジャー管理テーブル1800の空きエントリを見つけるには、SANアドレス1830が値-1であるエントリを探せばよい。

【００４６】ステップ2840で、新規追加されたデバイスがストレージ装置1060であると認識された場合には、ステップ2880で、ドライブ管理テーブル1400の空きエントリに、アクセス回数1430を0とし、キャッシュマネジャーSANアドレス1440を-1（独自のキャッシュを持つ場合）あるいは-2（独自キャッシュを持たない場合）として、更にそのデバイスのSANアドレスをSANアドレス1450に設定し、デバイス認識処理2800は終了する。なお、ドライブ管理テーブル1400の空きエントリは、SANアドレス1450が値-1であるエントリを探すことで見つけることができる。

【００４７】ステップ2850でキャッシュマネジャー1200でないと認識された場合には、ステップ2860でその他のデバイスの管理テーブルを設定するが、本発明には直接関係ないため設定する内容やテーブルについては説明を割愛する。そしてデバイス認識処理2800は終了する。

【００４８】なお、ステップ2850、ステップ2870、ステップ3000はキャッシュマネジャーマネジャー1030が実行する。

【００４９】図１４は、メタデータサーバー1020上のキャッシュマネジャーマネジャー1030が実行し、キャッシュマネジャー1200がキャッシュ機能を提供するストレージ装置1060の範囲を変更するキャッシュマネジャー構成変更処理3000のフローチャートである。キャッシュマネジャー構成変更処理3000はどのような契機で実行してもよいが、本実施形態ではデバイス認識処理2800で実行されるのに加え、三日に一度自動実行されるものとする。

【００５０】まずステップ3010で、SAN1040上にある全キャッシュマネジャー1200に対し、ストレージ装置1060への未反映データを全て反映するよう要求する。そしてステップ3020でキャッシュマネジャー1200から完了報告を受領し、ステップ3030で全キャッシュマネジャー1200から完了報告があったかどうかを判定する。なかったと判定された場合にはステップ3020にジャンプし、あったと判定された場合にはステップ3040で、ドライブ管理テーブル1400の全てのエントリのキャッシュマネジャーSANアドレス1440で値が-1（キャッシュ機能を有するストレージ装置1060であることを意味する）でないものに値-2（キャッシュ機能を提供されていないことを意味する）を代入する。そしてステップ3050でドライブ管理テーブル1400のアクセス回数1430を見て、キャッシュマネジャーSANアドレス1440が-2のエントリであって回数の多いものから順に、そのストレージ装置1060に対応するドライブ管理テーブル1400内のエントリ番号をドライブソートテーブル3200にセットしていき、そしてカレントソート番号3220に0を代入する。この時、対応するストレージ装置1060の無いエントリのSANアドレス1450には無効を示す値-1を入れる。そしてステップ3060でストレージ装置1060の割り当てが済んでいないキャッシュマネジャー1200がまだ存在するかどうかを判定し、なければキャッシュマネジャー構成変更処理3000は終了する。キャッシュマネジャー管理テーブル1800の最初のエントリから処理を始め、終端に達していなければ、キャッシュマネジャー1200がまだ残っている可能性があり、キャッシュマネジャー管理テーブル1800の終端に達したら処理終了である。あると判定された場合には、ステップ3070において、キャッシュマネジャー管理テーブル1800のエントリを上から順に一つ選択し、ステップ3080でそのSANアドレス1830が無効を示す-1であるかどうかを調べる。-1であればキャッシュマネジャー管理テーブル1800のエントリを指すポインタを次のエントリに進めた後にステップ3060にジャンプする。-1以外であれば、ステップ3090において後述するドライブ割り当て処理3400を行い、キャッシュマネジャー管理テーブル1800のエントリを指すポインタを次のエントリに進めた後に、ステップ3060にジャンプする。

【００５１】図１５は、キャッシュマネジャー構成変更処理3000の一部として実行されるドライブ割り当て処理3400のフローチャートである。

【００５２】まずステップ3410で選択しているキャッシュマネジャー1200に対応可能ドライブ数1820分のストレージ装置1060を割り当てたかどうかを判定し、割り当てていないと判定された場合にはステップ3420でカレントソート番号3220の示すドライブソートテーブル3200のエントリに入っている番号の示すストレージ装置1060を選択する。そしてステップ3430で、ステップ3420で選択したストレージ装置1060番号が有効であるかどうかを判定し、有効であればステップ3440でカレントソート番号3220の値に1を加え、ステップ3450で当該ストレージ装置1060に対応するドライブ管理テーブルエントリ1420のキャッシュマネジャーSANアドレス1440に当該キャッシュマネジャー1200の番号を入れることで、当該ストレージ装置1060に当該キャッシュマネジャー1200を割り当てる。そしてドライブ割り当て数のカウンタに１を加えた後に、ステップ3410にジャンプする。

【００５３】ステップ3410で割り当てたと判定された場合には、ステップ3460で当該キャッシュマネジャー1200に割り当てられているストレージ装置1060のSANアドレスを通知するコマンドを発行し、ドライブ割り当て処理3400は終了する。このコマンドは、SCSIのモードセレクトコマンドを用い、ベンダー固有のパラメーターページに当該キャッシュマネジャー1200がキャッシュ機能を提供するべきストレージ装置1060のSANアドレスをセットする。そしてドライブ割り当て処理3400は終了する。

【００５４】ステップ3430で無効なドライブ番号であると判定された場合には、ステップ3460にジャンプする。

【００５５】SAN1040にキャッシュマネジャー1200を新たに接続したときに、そのキャッシュマネジャーがキャッシュ機能を提供するストレージ装置1060を無しとして初期化しておき、その後そのキャッシュマネジャー1200を含めてキャッシュマネジャー構成変更処理3000を行うことで、システムを停止して特別なキャッシュマネジャー1200の追加処理を行うことを避けることができる。

【００５６】図１６は、キャッシュマネジャーマネジャー1030からキャッシュ機能を提供するストレージ装置1060を指定するMode Selectコマンドを受領したときに、キャッシュマネジャー1200が実行するキャッシュ割り当て処理3600のフローチャートである。

【００５７】まずステップ3610で、キャッシュマネジャー1200はコマンドを受領し、ステップ3620で、キャッシュドライブ管理テーブル3800のエントリに受領したストレージ装置1060のSANアドレスを全て代入して、キャッシュ割り当て処理3600は終了する。

【００５８】本実施形態では、キャッシュマネジャー1200へのストレージ装置1060の割り当てをキャッシュマネジャーマネジャー1030が行ったが、ストレージ装置1060のアクセス回数1430を参照してSAN1040の管理者が割り当てを決定するようにしてもよく、この場合も本発明を適用できることは言うまでもない。システムに複数のキャッシュマネジャー1200を設ける場合には、ストレージ装置1060のアクセス回数1430に応じて、アクセスが複数のキャッシュマネジャー1200に分散するように、キャッシュマネジャー1200にストレージ装置1060を割り当てると、１つのキャッシュマネジャー1200へのアクセスの集中が避けられる。

（２）第２の実施形態

第１の実施形態では、メタデータマネジャー1020が存在するSAN環境への本発明の適用を示した。第２の実施形態では、メタデータマネジャー1020が存在せず、SAN1040につながっているストレージ装置1060に格納されているファイルの位置情報を得る処理を、そのファイルを使用するコンピューター1010内のファイルシステムが行う場合について、本発明の適用を示す。なお、同じ説明の繰り返しを避けるため、第１の実施形態と違う部分についてのみ解説する。

【００５９】図１７は、本発明を適用するコンピューターシステムB1002の構成を示す。コンピューターシステムB1002にはメタデータサーバー1020が存在せず、代わりにコンピューター1010の中にファイル名からファイル格納位置の変換を行うファイルシステム1110とキャッシュマネジャーマネジャー1030が存在する。

【００６０】ドライブ管理テーブル1400、キャッシュマネジャーテーブル1800およびドライブソートテーブル3200はコンピューター1010上にあり、キャッシュマネジャーマネジャー1030が管理する。

【００６１】図１８は、コンピューター1010がSAN1040につながっているストレージ装置1060に対してアクセスする際の、ホストI/O処理A4000のフローチャートを示す。この処理は、ホストI/O処理2000とメタデータサーバー問い合わせ処理2200の組み合わせであり、対比をわかりやすくするために同様の処理には同じ番号をつけた。

【００６２】まずステップ4010でコンピューター1010上のアプリケーションプログラムがファイルアクセス要求を発行する。

【００６３】そしてステップ2220でそのパスの示すファイルを格納するストレージ装置1060のSANアドレスとその装置内のブロック番号からなる位置情報を求めて、ステップ2230でアクセス対象のストレージ装置1060に対応するドライブ管理テーブルエントリ1420のアクセス回数1430に1を加える。そして、同エントリ1420のキャッシュマネジャーSANアドレス1440を調べて、対象ストレージ装置1060にキャッシュ機能を提供しているキャッシュマネジャー1200があるかどうかの判定を行う。あると判定された場合には、ステップ2250で、以下の式を用いて位置情報の変換を行う。

【００６４】ブロック番号＝対象ストレージ装置のSANアドレス×1000000＋対象ブロック番号SANアドレス＝キャッシュマネジャーの持つSANアドレスこの変換により、アクセス対象となるSANアドレスはキャッシュマネジャー1200のSANアドレスとなり、ブロック番号には元々のアクセス対象のストレージ装置1060のSANアドレスとそのストレージ装置1060内ブロック番号を組み合わせた値が入ることになる。

【００６５】次に、求めた位置情報の示すストレージ装置1060に対し、格納ブロック番号のデータをアクセスするようI/O要求を発行し、ステップ2040で目的とするデータの送受信を行ってホストI/O処理2000は終了する。

【００６６】一方、ステップ2240で無いと判定された場合には、ステップ2030にジャンプする。

【００６７】なお、ステップ2230、ステップ2240およびステップ2250の処理はキャッシュマネジャーマネジャー1030が行う。

【００６８】デバイス認識処理2800及びキャッシュマネジャー構成変更処理3000はコンピューター1010上のキャ

ッシュマネジャーマネジャー1030が実行する。

【００６９】なお、コンピューター1010が複数ある場合には、キャッシュマネジャー構成変更処理3000を行うコンピューター1010は代表となる一台だけとし、それが終わったらその他のコンピューター1010に対して書き換えられたドライブ管理テーブル1400をLAN1100経由で渡す。代表以外のコンピューター1010は、受け取ったデータでドライブ管理テーブル1400を上書きする。このようにして、コンピューター1010間でドライブ管理テーブル1400の食い違いが起こることを防ぐ。

【００７０】なお単一のキャッシュマネジャー1200をメタデータサーバー1020に組み込んでもよい。その場合には、キャッシュマネジャーマネジャー1030はメタデータサーバー1020内でキャッシュマネジャー1200と通信し、キャッシュマネジャー1200はコンピューター1010からリード要求及びライト要求を受領して上記の通りその処理を行う。またこのようにキャッシュマネジャー1200を組み込んだメタデータサーバー1020を改めてキャッシュマネジャーと呼んでも構わない。

【００７１】第１及び第２の実施形態によれば、キャッシュマネジャー1200によってキャッシュ機能の提供を受けるストレージ装置1060のコンピューター1010から見た読み書き性能が向上する。一方キャッシュマネジャー1200がSAN1040に加わったことによって、キャッシュマネジャー1200とストレージ装置1060の間のネットワークトラフィック分が増加する。各キャッシュマネジャー1200の対応可能ドライブ数1820を適切に設定することによって、読み書き性能の向上とネットワークトラフィックの増加のトレードオフを図ることができる。

【００７２】

【発明の効果】以上述べたように本発明によれば、コンピュータ及び記憶装置が接続されるネットワークに、キャッシュマネジャーを接続するので、キャッシュ機能をもたない記憶装置に対してコンピュータ側の設定を変更せずにキャッシュ機能を提供することができ、コンピュータから見える記憶装置に対するデータの読み書き性能を向上させることができる。

【００７３】また本発明によれば、キャッシュ機能を提供する記憶装置を動的に変更することができるので、特にアクセス頻度の高い記憶装置に対して優先的にキャッシュ機能を提供するように動的な構成変更ができる。

【００７４】さらに本発明によれば、各キャッシュマネジャーについてキャッシュ機能を提供する記憶装置の台数の最大可能数を制限するので、ネットワークにキャッシュマネジャーが加わることによるトラフィック増を制限することができる。

【図面の簡単な説明】

【図１】本発明を適用する第１の実施形態のコンピューターシステムの構成図である。

【図２】第１の実施形態の他のコンピューターシステムの構成図である。

【図３】実施形態のキャッシュマネジャーの構成図である。

【図４】実施形態のドライブ管理テーブルのデータ構成図である。

【図５】実施形態のキャッシュマネジャー管理テーブルのデータ構成図である。

【図６】実施形態のドライブソートテーブルとカレントソート番号のデータ構成図である。

【図７】実施形態のキャッシュ管理テーブルとカレントキャッシュブロック番号のデータ構成図である。

【図８】実施形態のキャッシュドライブ管理テーブルのデータ構成図である。

【図９】第１の実施形態のホストI/O処理のフローチャートである。

【図１０】第１の実施形態のメタデータサーバー問い合わせ処理のフローチャートである。

【図１１】実施形態のキャッシュマネジャーリード処理のフローチャートである。

【図１２】実施形態のキャッシュマネジャーライト処理のフローチャートである。

【図１３】実施形態のデバイス認識処理のフローチャートである。

【図１４】実施形態のキャッシュマネジャー構成変更処理のフローチャートである。

【図１５】実施形態のドライブ割り当て処理のフローチャートである。

【図１６】実施形態のキャッシュ割り当て処理のフローチャートである。

【図１７】第２の実施形態のコンピューターシステムの構成図である。

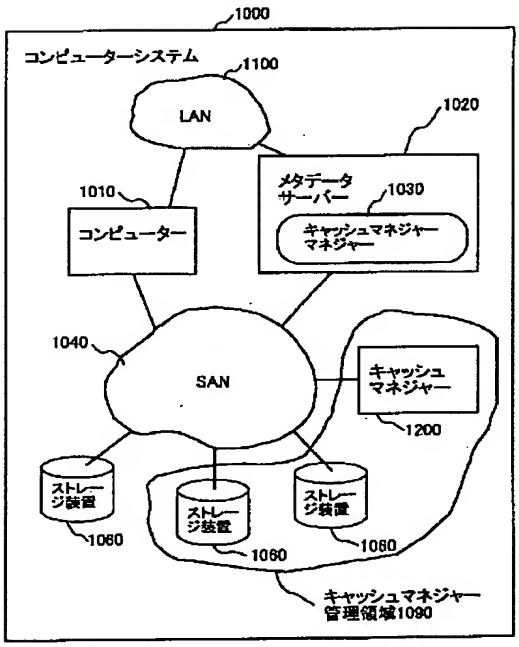
【図１８】第２の実施形態のホストI/O処理のフローチャートである。

【符号の説明】

1000…コンピューターシステム、1010…コンピューター、1020…メタデータサーバー、1030…キャッシュマネジャーマネジャー、1040…SAN、1200…キャッシュマネジャー、1090…キャッシュマネジャー管理領域、1100…LAN

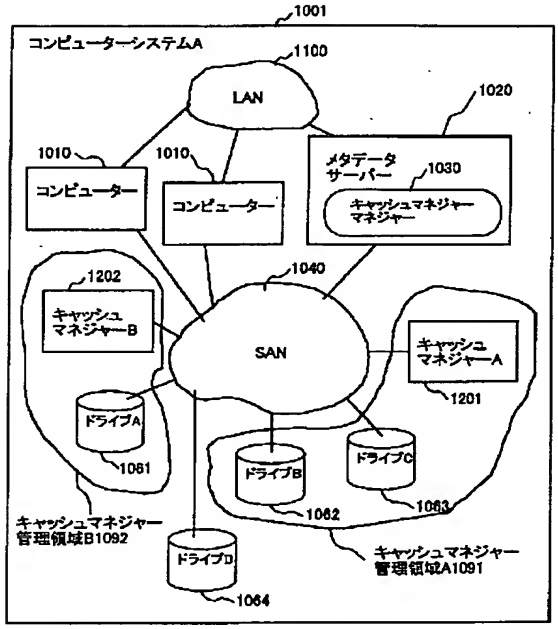
【図１】

図 1



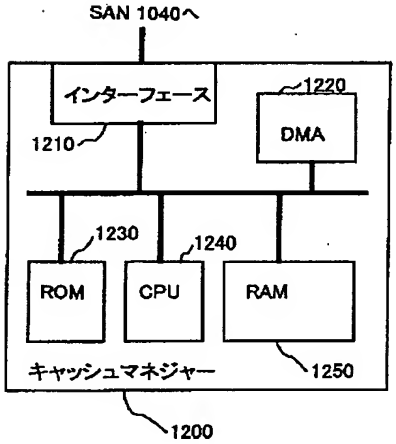
【図２】

図 2



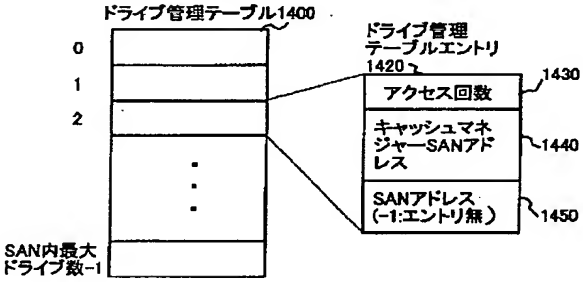
【図３】

図 3



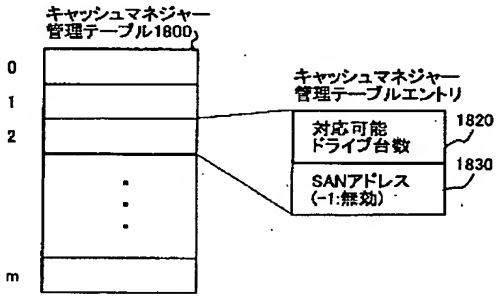
【図４】

図 4



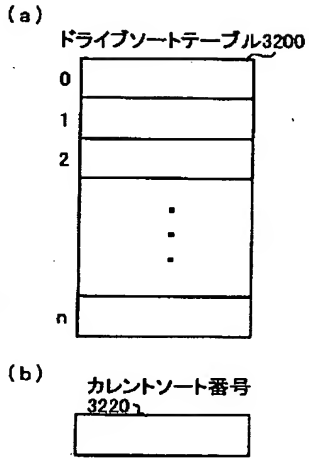
【図５】

図 5



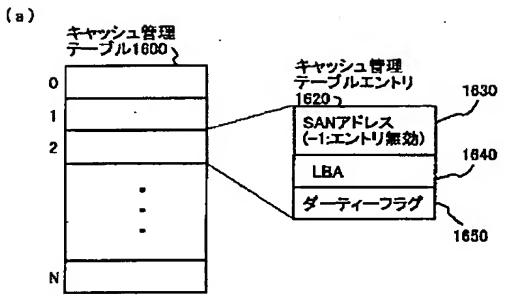
【図６】

図 6



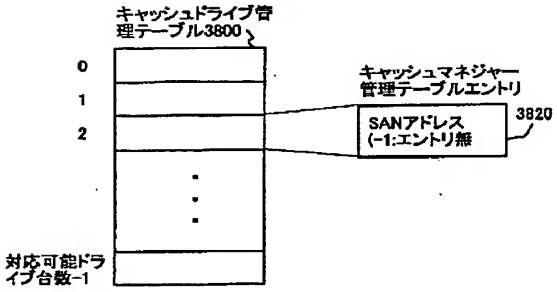
【図７】

図 7



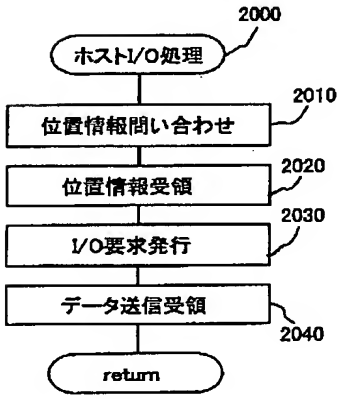
【図８】

図 8



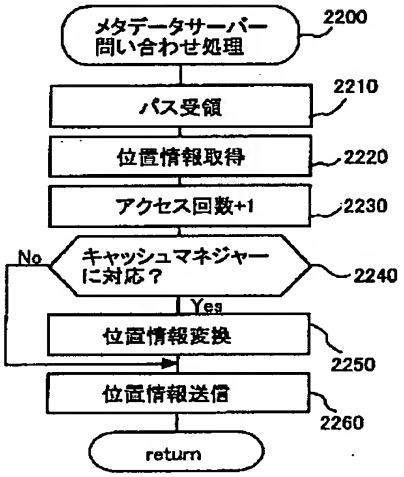
【図９】

図 9



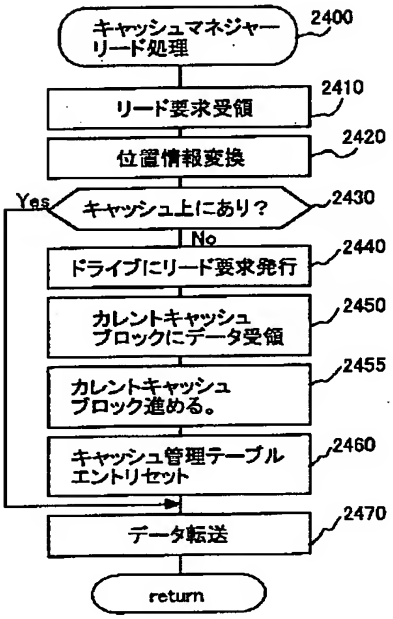
【図１０】

図 10



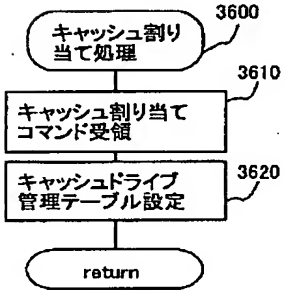
【図１１】

図 11



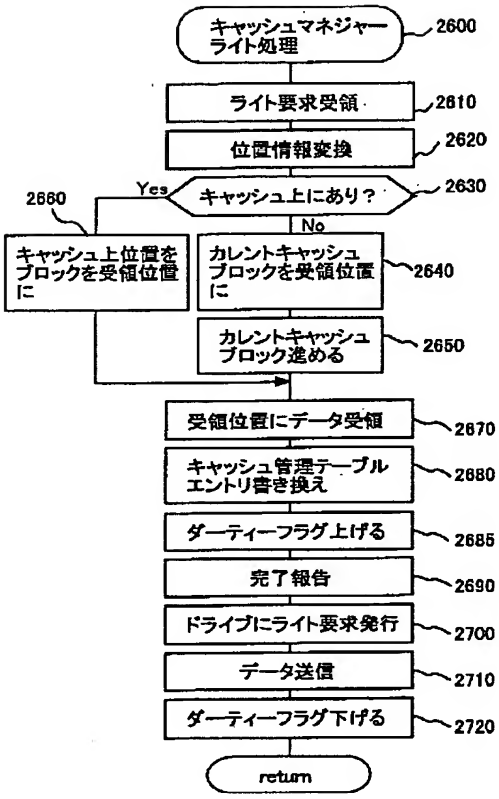
【図１６】

図 16



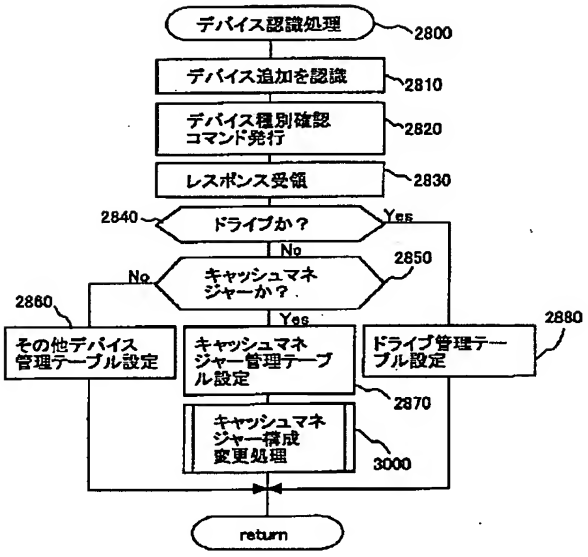
【図１２】

図 １２



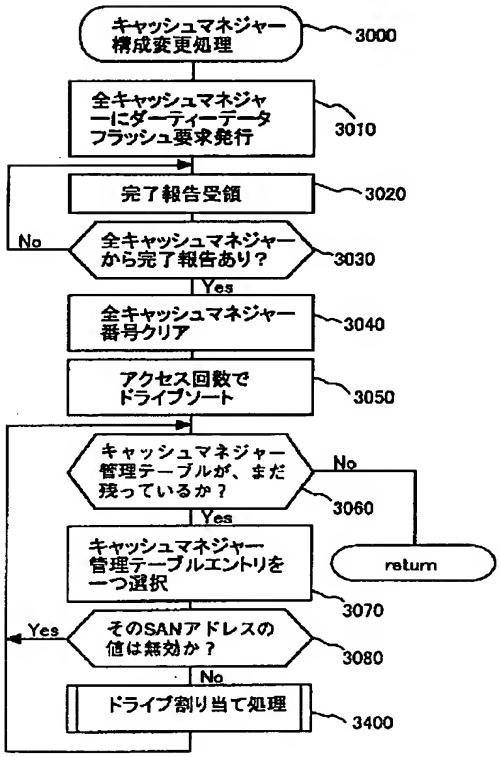
【図１３】

図 １３



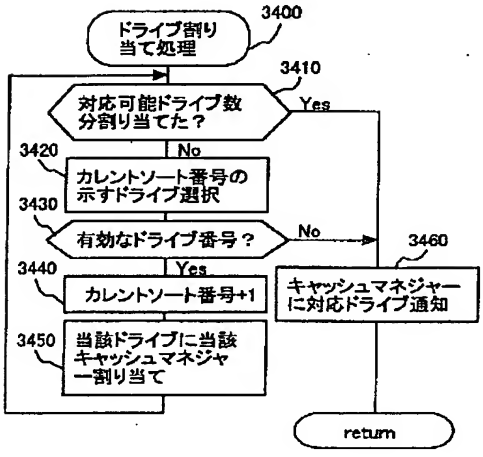
【図１４】

図 １４



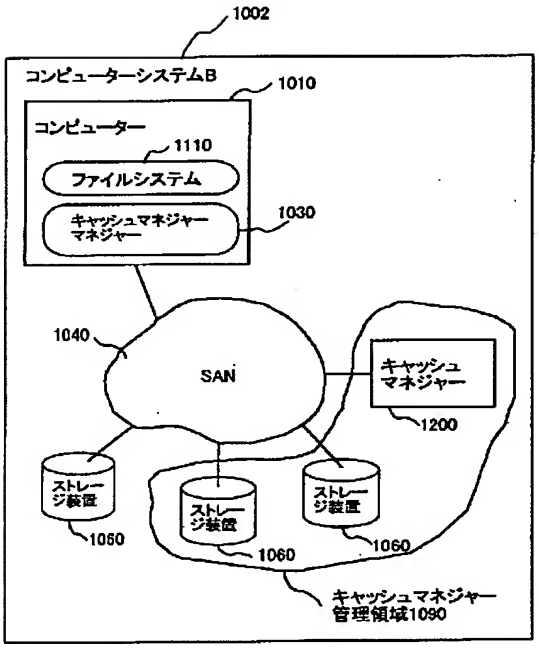
【図１５】

図 １５



【図１７】

図　１７



【図１８】

図　１８